

# MULTI-TIER VIDEO DELIVERY NETWORK

**Inventor(s):**

Thomas D. Miller, a citizen of United States, residing at,  
2501 Marks Road  
Alamo, CA 94507

Adityo Prakash, a citizen of India, residing at,  
600 Marlin Court  
Redwood Shores, CA 94065

Hon Hing So, a citizen of United States, residing at,  
1393 Meadow Ridge Circle  
San Jose, CA 95136

Kendall R. Haulotte, a citizen of United States, residing at,  
2090 Madera Court  
Tracy, CA 95376

Assignee:

**Pulsent Corporation**  
**1455 McCarthy Boulevard**  
**Milpitas, CA 95035**

Entity: Small business concern

## MULTI-TIER VIDEO DELIVERY NETWORK

### CROSS-REFERENCES TO RELATED APPLICATIONS

5 The present disclosure claims priority from U.S. Provisional Patent Application No. 60/224,748, filed August 10, 2000, and entitled "Multi-Tier Video Delivery Network," the disclosure of which is hereby incorporated by reference for all purposes.

### FIELD OF THE INVENTION

10 The present invention relates to broadband video content distribution networks utilizing Asynchronous Transfer Mode (ATM) or Internet Protocol (IP) backbone transport and a system of coordinated video servers, video storage devices, video caching devices and a video stream manager.

### BACKGROUND OF THE INVENTION

15 In the current environment, with the explosion in popularity of the Internet for entertainment, commerce and other new uses (as well as the old uses), many users expect to get large bandwidth data sets, such as the data that makes up a video sequence, transmitted over packet-based networks, one example of which is the Internet. A packet-  
20 based network is a network where the data is transmitted in a non-continuous fashion, which allows the physical connection to be shared by multiple users simultaneously. The end-users expect to get such video data in real, or near real, time. Real-time video streams consume considerable bandwidth, as the recipient expects to receive video at its frame rate. Thus, if the frame rate is thirty frames per second, the recipient expects the  
25 network to deliver, in one second or less, enough data to decode thirty frames of video.

For some transmission media, such as direct satellite broadcasting systems or cable television networks, sending real-time video is easy and has in fact been done for years. Unfortunately, such networks are only useful where only a few (many numbering in the hundreds) video streams are being sent and they are being sent simultaneously to  
30 the recipients, and cannot readily be adapted to sending different video streams at possibly different times, to many different places on the network.

One disadvantage of streaming content out from a centralized source every time that content is requested, even if it is done in non-real-time, is the cost of handling

full video connections to each of possibly millions of recipients. With high-bandwidth content and many recipients requesting a copy of the content, the network will slow under the weight of the traffic. One partial solution that attempts to overcome that disadvantage is to have the content streamed once from its source at a central location to many servers  
5 located nearer to recipients, thus reducing the number of copies of the data that have to flow over large portions of the network. While this solution would be useful if the content that is streamed to the network edge is used by recipients near each edge server, even more network bandwidth might be wasted as unused content is streamed from the central location to edge servers that might not ever serve up the content.

### SUMMARY OF THE INVENTION

The invention relates to a system and method of delivering video over a communication system comprising of a system of video servers, video storage devices,  
15 video caches and a broadband transmission network. Video material is initially received from content providers, encoded using a compression algorithm, and stored at a centralized location. Depending on the projected frequency of viewing of the content, it either remains stored solely at the centralized location for less frequently viewed content, or the projected frequently viewed content is proactively transmitted to storage systems  
20 located on the edge of the network close to end users. In addition to storage, certain video material is streamed live to the end user from the centralized location or other encoding facilities, via the local center. The transmission between the centralized facility and the local center is controlled by a stream manager capable, based upon various criteria, of prioritizing the transmission of different video streams both live and  
25 previously stored, determining whether certain video data should be stored at the local center, distributing the total volume of video data over multiple networks.

### BRIEF DESCRIPTION OF DRAWINGS

30 Fig. 1 illustrates the method and apparatus of playing video content archived at the centralized location

Fig. 2 illustrates the method and apparatus of playing live video content transmitted through the centralized location

Fig. 3A illustrates the method and apparatus of archiving video content at the local center

Fig. 3B illustrates the method and apparatus of playing video content archived at the local center

5

### DETAILED DESCRIPTION OF THE EMBODIMENTS

With reference to the exemplary drawings wherein like reference numerals indicate like or corresponding elements among the figures, embodiments of a system according to the present invention will now be described in detail.

10 Figure 1 illustrates the concept of the utilization of a multi tier system to deliver high quality video content from a centralized source to an end user. Customer A initiates a request for a specific file of video content. This request is sent to the local center 101 and through the local center router 102 to the local center video server 103. The local center video server 103 determines that the requested video content is located at  
15 the centralized center 104 and forwards a request through local center router 102 to the centralized center router 105, which sends the request to the centralized center video server 106. The location of the specific content is identified in the centralized center video storage 107 and played back to the local center via the centralized center video server 6 and router 105. The content is received at the local center via the local router  
20 102 and sent via the local video server 103 to the local center video cache 108.

After the local video cache has received 3-10 seconds of video content, it signals the home cache 109 that it is ready to transmit the content and after receiving an acknowledgement from the home cache, begins to transmit the video content. After the home cache has received 3-10 seconds of video content, it begins to play the material to  
25 the end user through the viewer of choice, either a personal computer 110 or a television 111. The home cache 109 and the local center video cache 108 maintain a flow of communication about the status of the fill level of the local center video cache 108 and ensures that a sufficient rate of content is transmitted to the home cache 109 to continuously enable an uninterrupted video stream to the end user.

30 The local center video cache 108 and the centralized center video server 106 also maintain a flow of communication about the status of the fill level of the local center video cache 108. In this case, however, the decision to transmit additional material is based both on the fill level of the cache and the available bandwidth of the transmission path 112 between the centralized center router 105 and the local center router 102. The

available bandwidth will be utilized as additional customers served out of the local center make requests for video content in the same manner as Customer A described above.

Although Fig. 1 depicts a system with four end users served out of a local center, it is envisioned that any number of customers may be served from a given center.

5 Furthermore, as the network expands in size, an additional tier(s) of serving centers may be inserted between the centralized center 104 and the local center 101.

Bandwidth prioritization is controlled by the stream manager 113 which dynamically allocates available bandwidth to individual video streams based on an algorithm that sends more content to the local center video cache when there is available  
10 bandwidth and less when the transmission path 112 is nearing its overall capacity.

As an individual video stream serving an end user from local center video cache 108 begins to reach the low end of its designated fill level, that stream is assigned a higher bandwidth priority when compared to caches for other video streams that are fuller. That priority is maintained until the cache supporting that individual stream is  
15 refilled to its threshold level. When overall demand for bandwidth on the transmission path 112 is light, the stream manager 113 will opt to fill the individual caches supporting requesting end users to capacities exceeding nominal fill levels in order to optimize the use of the dedicated transmission capacity. This will enable less bandwidth to be supplied to these caches in subsequent time increments when other end users make  
20 requests for video content or when the aggregate bandwidth for all streams is relatively high.

Although Fig. 1 depicts a single transmission path between the centralized location (center) 104 and the local center 101, it is contemplated that multiple physical transmission links can be implemented, depending on the overall engineered traffic  
25 volume. In the case of high volume of video data, the stream manager 113 also implements a load balancing algorithm to optionally distribute the total volume of traffic over the multiple transmission links, including transmission over the public Internet for lower priority traffic which has to be transmitted between the centralized center 104 and the local center 101. A similar load balancing algorithm can be implemented for  
30 distribution of traffic through multiple networks between the local center 101 and the home cache 109.

Utilization of the local center video cache 108 allows for the opportunity for the local center video server 103 to signal the video server 106 at the national center to retransmit data packets when they have been lost or corrupted during transmission.

Upon retransmission, lost packets can be inserted into the video cache 108 in the correct sequence prior to transmitting the information to the home cache 109. Similarly, the video decoder located in the television 110 (or a set top box adjunct) or in the personal computer 111, will have an opportunity to signal the local center video server 103 for retransmission when a packet has been lost or corrupted during transmission from the local center to the home. Upon retransmission, lost packets can be inserted into the home cache 109 in the correct sequence prior to transmitting the information to the computer or television of the end user.

Figure 2 illustrates the concept of the utilization of a multi tier system to deliver high quality live video material via a centralized source to end-users. Source material 201 is captured utilizing a video camera 202 and transmitted to the video encoder 204 at the centralized center. Alternatively, the encoder could be co-located at the recording site in order to conserve transmission bandwidth from the recording site to the centralized center. After the content is encoded, it is placed in a video cache 206 at the centralized location and streamed out to the local center via the video server 205 and the centralized center router 210 over a dedicated transmission path 211. The video content is received in the local center by the router 208 and directed by the video server 209 to the local center video cache 213.

If Customer A initiates a request for a specific file of video content, this request is sent to the local center and through the local center router 208 to the local center video server 209. The video server locates the content in the local center video cache 213 and begins to transmit the live video content from the video cache 213 to the home cache 214.

Unlike the previous example of playing back pre-recorded content, the cache sizes in the centralized, local and home caches must be kept relatively small, in the range of 1-10 seconds, in order to maintain the "live" characteristic of the broadcast. Also unlike the previous example, the live content is sent from the centralized center to the local center after Customer A requests it but should additional customers served from the local center request the content, it is sent out to those customers from the local center video server 209 through the local center router 208 using an existing process called multicasting without having to replicate bandwidth in transmission path 211 or resources in the centralized center.

Because of the time sensitive nature of the live material and the need to keep these queues small throughout the network, the stream manager 215 assigns a higher

priority to live streams when compared to the prerecorded streams in the previous example in order to assure that the caches supporting live material remain within their thresholds. Retransmission of lost or corrupted packets for live streams is assigned the highest priority by the stream manager 215.

Content Producers who are sending live video content will also have the option of recording and storing their content as it is being sent out live. In this case, the material will be simultaneously stored in the centralized location video storage 216 and made available for customer access as previously described in Figure 1.

Figure 3a illustrates the concept of proactively sending content from the centralized location 303 to the local centers 307 when there is a high probability that the content will be viewed by several or more of the end users served out of the local center. Source material 301 is produced and recorded 302 and transported to the centralized location 303 in a number of possible ways, including direct transmission, file transfer, physical delivery, etc., depending upon the preference of the content producer.

Upon receipt, the content is encoded 304 and routed via the video server 305 to a storage location 306. An availability schedule is negotiated with the content producer and an assessment is made between the content producer and the centralized location management as to whether the content is expected to be viewed frequently enough to warrant its proactive storage at one or more of the local centers.

If the decision is made to utilize local storage, the stream manager 311 is assigned the task of allocating network bandwidth on the transmission path 312. This task will normally be assigned a lower priority than the live and on demand pre-recorded content being transmitted over the transmission path 312 since it will not be as time sensitive. However, if bandwidth is available, the stream manager 311 will allocate additional bandwidth to the transmission of the video content designated for local archiving. If there is still available bandwidth after all of the content scheduled for local archiving has been delivered, the stream manager 311 will communicate with the content management software within the video server 305 and determine if additional content stored at the centralized location 306 would be a candidate for local archiving based on any one or a combination of the following parameters:

1. Available storage space in the local center video storage 310: The greater the amount of storage space available at the local center the greater the amount of additional content that will be archived at the local center and vice versa.

2. History of the content being previously accessed by end users served by that local center 307: The more frequently a video content is accessed by the end users, the higher the priority of the video content for local storage and vice versa.

3. History of similar content being previously accessed by end users served by that local center 307: The more frequently a video content is accessed by the end users, the higher the priority of similar video content for local storage and vice versa.

Figure 3b illustrates the concept of playing out the video content that is proactively archived at the local center. This is currently being performed by numerous video server suppliers and does not constitute a part of this invention. However, the overall system for integrating and efficiently managing the three types of video traffic (archived content at a centralized location, live content, and archiving and playing content from a local center) is an integral component of the invention.

It is important to note that within the communication system described above, the stream manager plays various important roles which include the following:

1. The stream manager prioritizes the transmission of the various streams based on a combination of the following:
  - i. The type of traffic which may be live or centrally stored.
  - ii. The amount of material remaining in the caches which require refreshing. The lower the remaining material in the caches the higher the priority of transmission.
  - iii. The speed at which the end user is viewing the material. If a particular video material is being viewed at a higher speed than other materials, the material will be given higher priority of transmission.
  - iv. The necessity to retransmit data due to corruption and lost packets.
2. The stream manager also determines the instantaneous amount of bandwidth required to transmit the video information to each end user and staggers the transmission of high bandwidth instants with lower bandwidth instants in other video data streams to produce a smoother, aggregated stream that is more efficient to transport over the video delivery network.



3. The stream manager determines which video content is a candidate for storage at the local facility based on the criteria described above.
4. The stream manager determines the load on the network and implements a load balancing algorithm to optionally distribute the total volume of traffic over
- 5 multiple networks including among others, the public Internet.

The above description is illustrative and not restrictive. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

10